

## **1. Evaluate / Specify Needs**

- 1.1. Define research needs, coverage & high-level concepts
- 1.2. Evaluate existing data & publications
  - 1.2.1. Determine if the data cover multiple time periods within the temporal coverage period desired
  - 1.2.2. Determine if the data are available at the geographic level of specificity required within the spatial coverage desired
  - 1.2.3. Determine the availability of the data (legal)
  - 1.2.4. Determine access and acquisition options
  - 1.2.5. Determine if the metadata available are sufficient for the needs of the project
  - 1.2.6. Determine relationship to overall mix of data within the project
  - 1.2.7. Classify the priority of the data to the project (critical, unique, richness, topical coherence, etc.)
  - 1.2.8. Identify costs involved in obtaining the data
  - 1.2.9. Is the proposed output structure for this data replicated elsewhere?
  - 1.2.10. Evaluate the quality of the data source in relation to the project output and subsequent research based on the project output.
- 1.3. Establish outputs & needed infrastructure
  - 1.3.1. Determine required and desirable output from the system
  - 1.3.2. Evaluate against existing system to identify new feature requirements
  - 1.3.3. Determine cost of developing new infrastructure
  - 1.3.4. Determine cost of implementing within current infrastructure
- 1.4. Identify specific concepts to be harmonized
  - 1.4.1. Identify concepts to harmonize
  - 1.4.2. Select concepts required or preferred from input data
- 1.5. Plan, create timetable, & identify needed infrastructure
- 1.6. Identify partners
  - 1.6.1. Identify potential partners and roles
  - 1.6.2. Negotiate required commitments
- 1.7. Prepare proposal and get funding
  - 1.7.1. Outline and assign proposal
  - 1.7.2. Write proposal
  - 1.7.3. Obtain institutional support
  - 1.7.4. Obtain required institutional approvals
  - 1.7.5. Submit for funding
  - 1.7.6. Respond to funder requests for alteration
  - 1.7.7. Obtain funding

## **2. Design / Redesign**

- 2.1. Identify sources
- 2.2. Design sampling methods
- 2.3. Design capture process
- 2.4. Specify data elements and related metadata
- 2.5. Specify processing / data cleaning methods
- 2.6. Specify evaluation plan
- 2.7. Organize research team

2.8. Design infrastructure

### 3. Build / Rebuild

- 3.1. Develop data capture processes
- 3.2. Create or enhance infrastructure components
- 3.3. Validate processes and tools
- 3.4. Test production systems
- 3.5. Finalize production systems

### 4. Collect

- 4.1. Select sources
- 4.2. Negotiate access and distribution rights
- 4.3. Capture data
  - 4.3.1. Receive data package (push by owner)
    - 4.3.1.1. Record receipt date and recipient
    - 4.3.1.2. Record media and format
    - 4.3.1.3. Record source (owner)
  - 4.3.2. Download data package (pull from provider)
    - 4.3.2.1. Record download date/time
    - 4.3.2.2. Record format and download URL
    - 4.3.2.3. Record source (Owner)
    - 4.3.2.4. Record actor who performed downloads (individual, machine, etc.)
  - 4.3.3. Download selected data (pull from provider)
    - 4.3.3.1. Record download date/time
    - 4.3.3.2. Record format and download URL
    - 4.3.3.3. Record filter used (selection criteria, database subset selected, etc.)
    - 4.3.3.4. Record selected variables
- 4.4. Obtain metadata
  - 4.4.1. Request provision of required metadata (from 2.4)
    - 4.4.1.1. Notify owner of requirements
  - 4.4.2. Search for required metadata from alternate sources (libraries, archives, repositories, researchers, etc.)
    - 4.4.2.1. Evaluate source options for validity/trustworthiness
    - 4.4.2.2. Clarify use/distribution rights
  - 4.4.3. Download metadata directly from owner (pull from provider)
    - 4.4.3.1. Record date/time and actor
    - 4.4.3.2. Record format and download URL
    - 4.4.3.3. Record filter used to locate
    - 4.4.3.4. Record selected subset (chapter, annex, pages, etc.)
  - 4.4.4. Process received metadata
    - 4.4.4.1. Record receipt date and recipient
    - 4.4.4.2. Record media and format
    - 4.4.4.3. Record source and distributor
    - 4.4.4.4. Create base bibliographic record
- 4.5. Create sample
  - 4.5.1. Determine appropriate sampling algorithm
  - 4.5.2. Apply algorithm to input data

4.5.3. Verify sample (distribution, coverage, etc.)

4.5.4. Verify case and unit weights

## **5. Process / Analyze**

5.1. Validate data against metadata

5.1.1. Visually verify that all variables are documented with name, definition, physical location, representation type, and coding specification

5.1.2. Run frequencies on all variables and compare against metadata checking for out of range values and undocumented codes

5.1.3. Run selected aggregations and compare to published statistics

5.1.4. Run selected aggregations and verify representative geographic coverage

5.2. Select and restructure data

5.2.1. Determine source and target structures

5.2.2. Map from source to target

5.2.3. Determine project related variables to add

5.2.4. Create target file(s)

5.3. Clean and anonymize data

5.3.1. Check for invalid values

5.3.2. Comparison with published aggregations

5.3.3. Identify outliers or unique cases for anonymization

5.3.4. Determine anonymization approach required

5.3.5. Apply anonymization algorithm to the data

5.4. Impute missing data

5.4.1. Determine imputation algorithm

5.4.2. Apply imputation algorithm

5.5. Harmonize selected data

5.5.1. Develop selection criteria for included variables

5.5.2. Select variables

5.5.3. Capture metadata on values including universe, definitions and instructions for question completion

5.5.4. Translate metadata

5.5.5. Identify relationships between values (within and between samples)

5.5.6. Establish harmonized response set with category, uniform definition and code

5.5.7. Map original categories and codes to harmonized categories

5.5.8. Define method for splitting and/or assigning categories based on complex code (other than one-to-one mapping)

5.5.9. Capture comparability between original and harmonized value

5.5.10. Capture comparability between samples from the same country over time

5.5.11. Complete recode script for coding harmonized value

5.5.12. Run recode script

5.5.13. Validate output

5.6. Calculate weights

5.7. Calculate aggregates

5.8. Validate processed data

5.9. Finalize data outputs

## **6. Archive / Preserve / Curate**

- 6.1. Ingest data & metadata
- 6.2. Enhance metadata
- 6.3. Capture process/provenance metadata
- 6.4. Preserve data & metadata
  - 6.4.1. Record receipt of all incoming data and metadata
  - 6.4.2. Scan metadata according to type
  - 6.4.3. Create bibliographic record for metadata object
  - 6.4.4. Create snapshot of data and metadata including: pre-specified data subsets, simple text codebook, DDI codebook, SAS, SPSS, and Stata setup files
  - 6.4.5. Validate DDI codebook: valid DDI, nonASCII-Latin correction
- 6.5. Undertake ongoing curation
  - 6.5.1. Draft DataCite record including: DOI, Title, Creators, Contributors, Publisher, Date, ObjectType, Abstract, Geographic Coverage, Funders, isNewVersionOf, isPartOf
  - 6.5.2. Verify draft DOI record with project manager
  - 6.5.3. File DOI with DataCite
  - 6.5.4. Verify with project manager when live project is ready for snapshot
  - 6.5.5. Request creation of landing page for previous version based on that version's DOI record
  - 6.5.6. Change previous version DOI updating link to landing page, adding isPreviousVersionOf information

## **7. Data Dissemination / Discovery**

- 7.1. Deploy release infrastructure
- 7.2. Preserve dissemination products
- 7.3. Deploy access control system / policies
- 7.4. Promote dissemination products
- 7.5. Provide data citation support
- 7.6. Enhance data discovery
- 7.7. Manager user support

## **8. Research / Publish**

- 8.1. Obtain listing of publications based on the data product
- 8.2. Maintain publication database
- 8.3. Manage versioning
- 8.4. Deposit metadata in related systems
  - 8.4.1. Enhance DDI content
  - 8.4.2. Adapt content for specified repository
  - 8.4.3. Set up for manual/automated harvesting; notify repository as required
- 8.5. Manage disclosure risk

## **9. Retrospective Evaluation**

- 9.1. Establish evaluation criteria
- 9.2. Gather evaluation inputs
- 9.3. Conduct evaluation
- 9.4. Determine future actions